2023 Issue 26 (2)

Thurstonian Scaling of Preference Ranking

Daniel M. Ennis

Background: The purpose of this report is to discuss an approach to converting preference ranks to intensities so that the scaled degree of hedonic difference among the ranked items can be determined. These preliminary results can then be used to predict paired preference and provide a basis for designing quantitative preference advertising claims or used to guide the development of products that are superior to competitors. Preference clams against competitors, especially if they are quantitative such as "two out of three people prefer…" are very powerful. To prepare for the test on which a claim will rely, it is often worthwhile to conduct preliminary or pilot research. Projections from this research can help to plan research aimed at the most compelling claims outcome.

Scenario: You have an interest in deriving hedonic scale values from ranked preference data for four beers. Three of these beers are your competitors and the fourth is one of your main brands. You would like to use a method that can efficiently and cost-effectively provide guidance on the design of a claims support test that you plan to conduct against one, or all, of your competitors. The final claim or claims will make count-based comparisons such as "2 out of 3 prefer…" against your competitors and you need preliminary guidance on what to expect before conducting the final claims test. You also plan to compare the derived values statistically for scaled differences as opposed to conducting non-parametric rank sum tests, which do not provide this information.

Alternative Methods: In order to place a set of items on a common hedonic scale, there are alternative approaches. One method is to take each pair of products and conduct a preference test on the six pairs and then concatenate the scale values. This is a laborious method, made especially difficult by the limitations imposed when testing alcoholic beverages. It is also highly time-consuming, thus driving up the cost of the central location test (CLT) that you are conducting if separate sessions are required. Another possible experimental method is to use the first-last choice method. This method was originally called MaxDiff, but this name refers to a different method which was first introduced in Psychophysics by Richardson[1,2] in 1938, so it is better to avoid this terminology. One of the reasons for the popularity of the first-last choice method is that scaling information can be obtained without using ratings. However, first-last choice ignores ranks for intermediate items which carry information. Preference ranking has advantages over multiple preference tests in cost and efficiency and measures and accounts for the intermediate values.

Thurstonian Scaling: Thurstonian scaling of fully ranked items, including intermediate values, has posed some computational challenges in the past and there is an extensive literature on how to handle them[3,4]. This problem was solved in the last few years with a computationally simple method[5]. Now it is possible to determine Thurstonian d′ values and their variances obtained from complete or partial ranking of items such as products or features. When multiple items are evaluated, it is more efficient than using multiple paired testing. Partial ranking, where subsets of three or four items are evaluated provides an opportunity to evaluate many alternatives without imposing a severe task load on each respondent.

Preference Ranking: You recruit 300 consumers in the premium light beer category and obtain each respondent's preference rankings on the four beers using a balanced presentation order that takes into account product positions, product sequences, and sequence spread as discussed in previous technical reports[6]. Your analysis plan is to use the Thurstonian model for rank data available in IFPrograms. This model provides d′ values for each product along with their sample estimate variances. Statistical comparisons are made using differences in the d′ values.

Table 1 shows a partial set of preference rankings for the four products from least preferred to most preferred. During the test, each product is assigned a 3-digit random number and preference ranked. In the table, product 4 refers to your product and the other three are competitors. There are no ties.

Least	…	…	Most
4	2	1	3
1	3	2	4
3	1	4	2
2	4	1	3
…	…	…	…
3	1	2	4
4	3	1	2
1	4	2	3
3	2	4	1

Table 1. A subset of the preference ranking data from least preferred to most preferred.

Thurstonian Ranking Analysis: Table 2 shows the Thurstonian d′ values and their variances. These values were obtained from IFPrograms using the Thurstonian model for rank data. This table shows the superiority of your brand to your competitors, where Competitor 1 is assigned a zero and the other products, including yours, are scaled relative to this product. Table 3 shows a difference comparison between your product and the three competitors. Your product has the highest hedonic value followed by Competitor 3. Competitor 1 has the lowest hedonic value. According to the hypothesis test, competitor 3 is not significantly different at 5% level from your product and the other two products are demonstratively less preferred.

Predicted Preference: From the d′ values obtained in the pilot preference ranking experiment, paired preference can be predicted. Figure 1 and Table 4 shows these results. Based on

these predictions you expect that on average your product would be preferred by about a 2 out of 3 margin to Competitor 1 and your product would be preferred by a little more than a 3 out of 5 margin to Competitor 2. Your product and Competitor 3 are more comparable than the other two, although the predicted preference favors your brand. This information is valuable to determine the sample size and test power when designing your test to support a persuasive claim.

Power and Sample Sizes for the Claims Tests: In order to determine the sample sizes need to conduct the planned claim support tests, you will need to specify an alpha level (α), an alternative hypothesis, and a power value. If the size of the differences detected in the ranking experiment are accurate and they represent the real values, you do not want to miss those differences often, so you consider power values of 95% and 99%. Using the generally accepted value of 0.05 as the α level in claims support, you can now proceed to determine the sample sizes recommended. Table 5 shows the sample sizes needed to declare a significant effect at 95% and 99% power to support finding the preferences reported in Table 4 with α = 0.05.

Conclusion: From this preliminary research, you decide to conduct the preference claims tests, with 300 participants, only against Competitors 1 and 2. Although this sample size is not needed for the comparison to Competitor 1, it may be required for network clearance and population representation. You decide against recommending a comparison to Competitor 3 on the grounds that the difference may be too small to be consumer-relevant and would require a large sample size. Preference ranking, attribute ranking of sensory variables, or ranking of product features on importance to purchase intent, can be a useful tool to determine the scale values using a Thurstonian model. This type of analysis was not possible without substantial computational complexity or the use of approximations a decade ago. Now it is possible to rapidly obtain these scale values using a very efficient method that applies to complete or partial ranks.

Product	d′	Variance
Competitor 1	0	0
Competitor 2	0.154	0.004
Competitor 3	0.332	0.022
Own Brand	0.605	0.004

Table 2. Thurstonian d′ values and variances. Competitor 1 is assigned a zero value.

Comparison	d′	p-Value
To Competitor 1	0.605	0.001
To Competitor 2	0.451	0.001
To Competitor 3	0.272	0.099

Table 3. d′ differences between your brand and the three competitors.

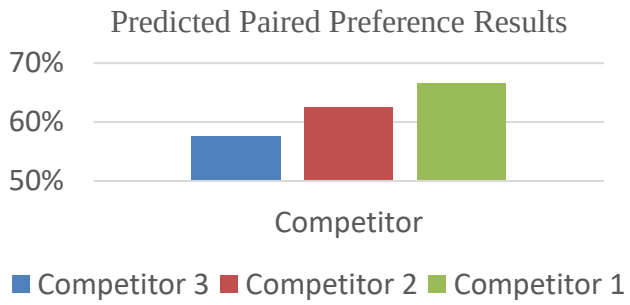

Figure 1. Predicted paired preference results based on preference ranking.

Comparison	Predicted Preference
To Competitor 1	66.6%
To Competitor 2	62.5%
To Competitor 3	57.6%

Table 4. Predicted preference results in a head-to-head comparison of your brand and each competitor.

Comparison	Power: 95%	Power: 99%
To Competitor 1	102	145
To Competitor 2	182	258
To Competitor 3	484	693

Table 5. Sample size requirements to detect Table 4 differences.

References

1. Torgerson, W. S. (1958). *Theory and Methods of Scaling*. New York: Wiley.
2. Ennis, D. M., Mullen, K. and Fritjers, J. E. R. (1988). Variants of the method of triads: Unidimensional Thurstonian models, *British Journal of Mathematical and Statistical Psychology*, 41, 25-36.
3. Böckenholt, U. (1992). Thurstonian models for partial ranking data, *British Journal of Mathematical and Statistical Psychology*, 43, 31-49.
4. Yao, G., and Böckenholt, U. (1999). Bayesian estimation of Thurstonian ranking models based on the Gibbs sampler, *British Journal of Mathematical and Statistical Psychology*, 52, 79-92.
5. Ennis, D. M. & Ennis, J. M. (2013). A Thurstonian ranking model with rank-induced dependencies. *Journal of Classification*. **30**, 124-147
6. Ennis, D.M. and Rousseau, B. (Eds.) (2020). *Tools and Applications of Sensory and Consumer Science*. pp. 158-159, 164-167. The Institute for Perception: Richmond, Va.